IN THE PATENT AND TRADEMARK OFFICE

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as
[ ] First Class mail
[ ] Express Mail Post Office to Addressee
in an envelope addressed to:
Mail Stop Patent Application
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

(Express Mail Tracking Number)
EV 032217058 US

on: 10/02/2003 by: Melissa Patanjia

Signature: Melissa Patanjia

APPLICATION FOR LETTERS PATENT

for

Methods and Compositions For Producing Secreted Trimeric Receptor Analogs And Biologically Active Fusion Proteins

Inventor:
Peng Liang of Nashville, TN

Assignee:
GenHunter Corporation of Nashville, TN

Methods and Compositions For Producing Secreted Trimeric Receptor Analogs And Biologically Active Fusion Proteins

Inventor: Peng Liang

Assignee: GenHunter Corporation

FIELD OF THE INVENTION

The present invention relates to methods for protein expression, and more specifically, for creating and expressing secreted and biologically active trimeric proteins, such as trimeric soluble receptors.

BACKGROUND OF INVENTION

In multicellular organisms, such as humans, cells communicate with each other by the so-called signal transduction pathway, in which a secreted ligand (e.g. cytokines, growth factors or hormones) binds to its cell surface receptor(s), leading to receptor activation. The receptors are membrane proteins, which consist of an extracellular domain responsible for ligand binding, a central transmembrane region followed by a cytoplasmic domain responsible for sending the signal downstream. Signal transdcution can take place in the following three ways: paracrine (communication between neighboring cells), autocrine (cell communication to itself) and endocrine (communication between distant cells through circulation), depending on the source of a secreted signal and the location of target cell expressing a receptor(s). One of the general mechanisms underlying receptor activation, which sets off a cascade of events beneath the cell membrane including the activation of gene expression, is that a polypeptide ligand such as a cytokine, is present in an oligomeric form, such as a homo-dimmer or

trimer, which when bound to its monomeric receptor at the cell outer surface, leads to the oligomerization of the receptor. Signal transduction pathways play a key role in normal cell development and differentiation, as well as in response to external insults such as bacterial and viral infections. Abnormalities in such signal transduction pathways, in the form of either underactivation (e.g. lack of ligand) or overactivation (e.g. too much ligand), are the underlying causes for pathological conditions and diseases such as arthritis, cancer, AIDS, and diabetes.

One of the current strategies for treating these debilitating diseases involves the use of receptor decoys, such as soluble receptors consisting of only the extracellular ligand-binding domain, to intercept a ligand and thus overcome the overactivation of a receptor. The best example of this strategy is the creation of Enbrel, a dimeric soluble TNF-α receptor-immunoglobulin (IgG) fusion protein by Immunex (Mohler et al., 1993; Jacobs et al., 1997), which is now part of Amgen. The TNF family of cytokines is one of the major pro-inflammatory signals produced by the body in response to infection or tissue injury. However, abnormal production of these cytokines, for example, in the absence of infection or tissue injury, has been shown to be one of the underlying causes for diseases such as arthritis and psoriasis. Naturally, a TNF–α receptor is present in monomeric form on the cell surface before binding to its ligand, TNF-α, □□□□□ exists, in contrast, as a homotrimer (Locksley et al., 2001). Accordingly, fusing a soluble TNF–α receptor with the Fc region of immunoglobulin G1, which is capable of spontaneous dimerization via disulfide bonds (Sledziewski et al., 1992 and 1998), allowed the secretion of a dimeric soluble TNF–α receptor (Mohler et al., 1993; Jacobs et al., 1997). In comparison with the monomeric soluble receptor, the dimeric TNF–α

receptor II –Fc fusion has a greatly increased affinity to the homo-trimeric ligand. This provides a molecular basis for its clinical use in treating rheumatoid arthritis (RA), an autoimmune disease in which constitutively elevated TNF-α, a major pro-inflammatory cytokine, plays an important causal role. Although Enbrel was shown to have a Ki in the pM range (ng/mL) to TNF-α (Mohler et al., 1993), 25 mg twice a week subcutaneous injections, which translates to μg/mL level of the soluble receptor, are required for the RA patients to achieve clinical benefits (www.enbrel.com). The high level of recurrent Enbrel consumption per RA patients has created a great pressure as well as high cost for the drug supply, which limits the accessibility of the drug to millions of potential patients in this country alone.

In addition to the TNF-α family of potent proinflmmatroy cytokines, the HIV virus that causes AIDS also uses a homo-trimeric coat protein, gp120, to gain entry into CD-4 positive T helper cells in our body (Kwong et al., 1998). One of the earliest events during HIV infection involves the binding of gp120 to its receptor CD-4, uniquely expressed on the cell surface of T helper cells (Clapham et al., 2001). Monomeric soluble CD-4 was shown over a decade ago as a potent agent against HIV infection (Clapham et al., 1989) however, the excitement was sadly dashed when its potency was shown to be limited only to laboratory HIV isolates (Daar et al., 1990). It turned out that HIV strains from AIDS patients, unlike the laboratory isolates, had a much lower affinity to the monomeric soluble CD-4, likely due to the sequence variation on the gp120 (Daar et al., 1990). Although the dimeric soluble CD-4-Fc fusion proteins have been made, these decoy CD-4 HIV receptors showed little antiviral effect against natural occurring

HIVs from AIDS patients, both in the laboratories and in clinics, due to the low affinity to the gp120 (Daar et al., 1990).

Clearly, there is a great need to be able to create secreted homo-trimeric soluble receptors or biologically active proteins, which can have perfectly docked binding sites, hence higher affinity, to their naturally occurring homo-trimeric ligands, such as the TNF family of cytokines and HIV coat proteins. Such trimeric receptor decoys theoretically should have a much higher affinity than its dimeric counterparts to their trimeric ligand. Such rationally designed soluble trimeric receptor analogs could significantly increase the clinical benefits as well as lower the amount or frequency of the drug injections for each patient. To be therapeutically feasible, like immunoglobulin Fc, the desired trimerizing protein moiety should ideally be part of a naturally secreted protein that is both abundant in the body and capable of efficient self-trimerization.

Collagen is a family of fibrous proteins that are the major components of the extracellular matrix. It is the most abundant protein in mammals, constituting nearly 25% of the total protein in the body. Collagen plays a major structural role in the formation of bone, tendon, skin, cornea, cartilage, blood vessels, and teeth (Stryer, 1988). The fibrillar types of collagen I, II, III, IV, V, and XI are all synthesized as larger trimeric precursors, called procollagens, in which the central uninterrupted triple-helical domain consisting of hundreds of "G-X-Y" repeats (or glycine repeats) is flanked by non-collagenous domains (NC), the N- propeptide and the C-propeptide (Stryer, 1988). Both the C- and N-terminal extensions are processed proteolytically upon secretion of the procallagen, an event that triggers the assembly of the mature protein into collagen fibrils which forms an insoluble cell matrix (Prockop et al., 1998). The shed trimeric C-

propeptide of type I collagen is found in the blood of normal people at a concentration in the range of 100-600 ng/mL, with children having a higher level which is indicative with active bone formation.

Type I, IV, V and XI collagens are mainly assembled into heterotrimeric forms consisting of either two α-1chains and one α-2 chain (for Type I, IV, V), or three different α chains (for Type XI), which are highly homologous in sequence. The type II and III collagens are both homotrimers of α-1chain. For type I collagen, the most abundant form of collagen, stable α-1(I) homotrimer is also formed and is present at variable levels (Alvares et al., 1999) in different tissues. Most of these collagen C-propeptide chains can self-assemble into homotrimers, when over-expressed alone in a cell. Although the N-propeptide domains are synthesized first, molecular assembly into trimeric collagen begins with the in-register association of the C-propeptides. It is believed the C-propeptide complex is stabilized by the formation of interchain disulfide bonds, but the necessity of disulfide bond formation for proper chain registration is not clear. The triple helix of the glycine repeats and is then propagated from the associated C-termini to the N-termini in a zipper-like manner. This knowledge has led to the creation of non-natural types of collagen matrix by swapping the C-propetides of different collagen chains using recombinant DNA technology (Bulleid et al., 2001). Non-collagenous proteins, such as cytokines and growth factors, also have been fused to the N-termini of either pro-collagens or mature collagens to allow new collagen matrix formation, which is intended to allow slow release of the noncollagenous proteins from the cell matrix (Tomita et al., 2001). However, under both circumstances, the C-

propeptides are required to be cleaved before recombinant collagen fibril assembly into an insoluble cell matrix.

SUMMARY OF THE INVENTION

Disclosed here is an invention that allows any soluble receptors or biologically active polypeptides to be made into trimeric forms as secreted proteins. The essence of the invention is to fuse any soluble receptors and biologically active proteins in-frame to the C-propeptide domain of fibrillar collagen, which is capable of self-trimerization, using recombinant DNA technology. The resulting fusion proteins when expressed in eukaryotic cells are secreted as soluble proteins essentially all in trimeric forms covalently strengthened by inter-molecular disulfide bonds formed among three C-propeptides.

In one aspect of the invention, a method for producing secreted trimeric fusion proteins is disclosed, comprising the following: (a) introducing into a eukaryotic host cell a DNA construct comprising a promoter which drives the transcription of an open reading frame consisting of a signal peptide sequence which is linked in-frame to a non-collagen polypeptide to be trimerized, which in turn is joined in-frame to the C-terminal portion of collagen capable of self-trimerization; (b) growing the host cell in an appropriate growth medium under physiological conditions to allow the secretion of a trimerized fusion protein encoded by said DNA sequence; and (c) isolating the secreted trimeric fusion protein from a host cell.

Within one embodiment, the signal peptide sequence is the native sequence of the protein to be trimerized. Within another embodiment, the signal peptide sequence is from a secreted protein different from that to be trimerized. Within one embodiment, the

non-collagen polypeptide to be trimerized is a soluble receptor consisting of the ligand binding domain(s). Within one embodiment, the C-terminal portion of collagen is the C-propeptide without any triple helical region of collagen (Sequence IDs: 3-4). Within another embodiment, the C-terminal collagen consists of a portion of the triple helical region of collagen as linker to the non-collagenous proteins to be trimerized (Sequence IDs: 1-2). Within another embodiment, the C-terminal portion of collagen has a mutated or deleted BMP-1 protease recognition site (Sequence IDs: 3-4).

In one aspect of the invention, a method for producing a secreted trimeric fusion protein is disclosed, comprising the following: (a) introducing into a eukaryotic host cell a DNA construct comprising a promoter which drives the transcription of an open reading frame consisting of a signal peptide sequence which is linked in-frame to a non-collagen polypeptide to be trimerized, which in turn is joined in-frame to the C-terminal portion of collagen capable of self-trimerization, selected from pro.alpha.1(I), pro.alpha.2(I), pro.alpha.1(II), pro.alpha.1(III), pro.alpha.1(V), pro.alpha.2(V), pro.alpha.1(XI), pro.alpha.2(XI) and pro.alpha.3(XI); (b) growing the host cell in an appropriate growth medium under physiological conditions to allow the secretion of a trimerized fusion protein encoded by said DNA sequence; and (c) isolating the secreted trimeric fusion protein from a host cell.

In a preferred embodiment, the non-collagen polypeptide to be trimerized is the soluble TNF-RII (p75) (Sequence IDs: 9-12). In another preferred embodiment, the non-collagen polypeptide to be trimerized is soluble CD-4, the co-receptor of HIV (Sequence IDs: 13-16). In yet another preferred embodiment, the non-collagen polypeptide to be trimerized is a placental secreted alkaline phosphatase (Sequence IDs: 5-8).

In one aspect of the invention, a method for producing a secreted trimeric fusion protein is disclosed, comprising the following: (a) introducing into a eukaryotic host cell a first DNA construct comprising a promoter which drives the transcription of an open reading frame consisting of a signal peptide sequence which is linked in-frame to a non-collagen polypeptide to be trimerized, which in turn is joined in-frame to the C-terminal portion of collagen capable of self-trimerization, selected from pro.alpha.1(I), pro.alpha.2(I), pro.alpha.1(II), pro.alpha.1(III), pro.alpha.1(V), pro.alpha.2(V), pro.alpha.1(XI), pro.alpha.2(XI) and pro.alpha.3(XI); (b) introducing into a eukaryotic host cell a second DNA construct comprising a promoter which drives the transcription of an open reading frame consisting of a second signal peptide sequence which is linked in-frame to a second non-collagen polypeptide to be trimerized, which in turn is joined in-frame to the second C-terminal portion of collagen capable of self-trimerization, selected from pro.alpha.1(I), pro.alpha.2(I), pro.alpha.1(II), pro.alpha.1(III), pro.alpha.1(V), pro.alpha.2(V), pro.alpha.1(XI), pro.alpha.2(XI) and pro.alpha.3(XI); (c) growing the host cell in an appropriate growth medium under physiological conditions to allow the secretion of a trimerized fusion protein encoded by said first and second DNA sequences; and (d) isolating the secreted trimeric fusion protein from the host cell.

In one aspect of the invention, a method for producing a secreted trimeric fusion protein is disclosed, comprising the following: (a) introducing into a eukaryotic host cell a first DNA construct comprising a promoter which drives the transcription of an open reading frame consisting of a signal peptide sequence which is linked in-frame to a non-collagen polypeptide to be trimerized, which in turn is joined in-frame to the C-terminal portion of collagen capable of self-trimerization, selected from pro.alpha.1(I),

pro.alpha.2(I), pro.alpha.1(II), pro.alpha.1(III), pro.alpha.1(V), pro.alpha.2(V), pro.alpha.1(XI), pro.alpha.2(XI) and pro.alpha.3(XI); (b) introducing into a eukaryotic host cell a second DNA construct comprising a promoter which drives the transcription of an open reading frame consisting of a second signal peptide sequence which is linked in-frame to a second non-collagen polypeptide to be trimerized, which in turn is joined in-frame to a second C-terminal portion of collagen capable of self-trimerization, selected from pro.alpha.1(I), pro.alpha.2(I), pro.alpha.1(II), pro.alpha.1(III), pro.alpha.1(V), pro.alpha.2(V), pro.alpha.1(XI), pro.alpha.2(XI) and pro.alpha.3(XI); (c) introducing into a eukaryotic host cell a third DNA construct comprising a promoter which drives the transcription of an open reading frame consisting of a third signal peptide sequence which is linked in-frame to a third non-collagen polypeptide to be trimerized, which in turn is joined in-frame to a third C-terminal portion of collagen capable of self-trimerization, selected from pro.alpha.1(I), pro.alpha.2(I), pro.alpha.1(II), pro.alpha.1(III), pro.alpha.1(V), pro.alpha.2(V), pro.alpha.1(XI), pro.alpha.2(XI) and pro.alpha.3(XI); (d) growing the host cell in an appropriate growth medium under physiological conditions to allow the secretion of a trimerized fusion protein encoded by said first and second DNA sequences; and (e) isolating the secreted trimeric fusion protein from the host cell.

The following are the advantages of this invention: (1) collagen is the most abundant protein secreted in the body of a mammal, constituting nearly 25% of the total proteins in the body; (2) the major forms of collagen naturally occur as trimeric helixes, with their globular C-propeptides being responsible for the initiating of trimerization; (3) the trimeric C-propeptide of collagen proteolytically released from the mature collagen is

found naturally at sub microgram/mL level in the blood of mammals and is not known to be toxic to the body; (4) the linear triple helical region of collagen can be included as a linker with predicted 2.9 Å spacing per residue, or excluded as part of the fusion protein so the distance between a protein to be trimerized and the C-propeptide of collagen can be precisely adjusted to achieve an optimal biological activity; (5) the recognition site of BMP1 which cleaves the C-propeptide off the pro-collagen can be mutated or deleted to prevent the disruption of a trimeric fusion protein; (6) the C-proptide domain provides a universal affinity tag, which can be used for purification of any secreted fusion proteins created by this invention.

In contrast to the Fc Tag technology (Sledziewski et al., 1992 and 1998), with which secreted dimeric fusion proteins can be created, this timely invention disclosed herein enables the creation and secretion of soluble trimeric fusion proteins for the first time. Given the fact that a homotrimer has 3-fold symmetry, whereas a homodimmer has only 2-fold symmetry, the two distinct structural forms theoretically can never be perfectly overlaid (Fig 1). As such, neither the homodimeric soluble TNF-R-Fc (e.g. Enbrel), nor the soluble CD4-Fc fusion proteins, could have had an optimal interface for binding to their corresponding homotrimeric ligands, TNF-α and HIV gp120, respectively. In contrast, homotrimeric soluble TNF receptors and CD4 created by the current invention are trivalent and structurally have the potential to perfectly dock to the corresponding homotrimeric ligands. Thus, these trimeric soluble receptor anologs can be much more effective in neutralizing the biological activities of their trimeric ligands. With this timely invention, more effective yet less expensive drugs, such as trimeric soluble TNF-R and CD4 described in the preferred embodiments, can be readily and

rationally designed to combat debilitating diseases such as arthritis and AIDS. Trimeric soluble gp120 can also be created with this invention, which could better mimic the native trimeric gp120 coat protein complex found on HIV viruses, and used as a more effective vaccine compared to non-trimeric gp120 antigens previously used. Also chimeric antibodies in trimeric form can be created with the current invention, which could endow greatly increased avidity of an antibody in neutralizing its antigen.

BRIEF DESCRIPTION OF DRAWINGS AND SEQUENCE LISTINGS

Fig.1 is a schematic representation of the method according to the invention compared to prior dimeric immunoglobulin Fc fusion.

On the left: Structural characteristics of a homodimeric soluble sTNF RII receptor-Fc fusion, such as Amgen's Enbrel, in either ligand-free or -bound form as indicated.

Domains labeled in green denote soluble TNF-RII. Note that the Fc (labeled in light blue with inter-chain disultfie bonds in red) fusion protein is dimeric in structure. Given its 2-fold symmetry, the dimeric Fc fusion protein is bivalent and thus theoretically does not have the optimal conformation to bind to a homotrimeric ligand, such as TNF-α (labeled in brown), which has a 3-fold symmetry.

On the right: Structural characteristics of a trimeric soluble sTNF RII receptor-C-propeptide fusion.

Given its 3-fold symmetry, a sTNF RII-Trimer fusion protein is trivalent in nature, thus can perfectly dock to its trimeric ligand TNF-α. □□□C-propeptide of collagen capable of self trimerization is labeled in dark blue with inter-chain disulfide bonds labeled in red.

Fig.2 illustrates the structures of pTRIMER plasmid vectors for creating secreted trimeric fusion proteins. Any soluble receptor- or biological active polypeptide-encoding cDNAs can be cloned into the unique Hind III or Bgl II sites to allow in-frame fusion at the C-termini to the α (I) collagen containing C-propetide sequence for trimerization. The pTRIMER(T0) construct contains part of the glycine-repeats (GXY)n upstream of the C-propeptide; whereas the pTRIMER(T2) contains only the C-propeptide domain with a mutated BMP-1 protease recognition site.

Fig. 3 illustrates the expression and secretion of disulfide bond-linked trimeric collagen fusion proteins.

A. Western blot analysis of the trimerization of human placental alkaline phosphatase (AP) when fused to the C-propeptides of α(I) collagen. The expression vectors encoding either AP alone or AP-C-propeptide fusions in pTRIMER vectors were transiently transfected into HEK293T cells. Forty-eight hours later, the conditioned media (20 μL) of each transfected cells as indicated were boiled for 5 minutes in equal volume of 2X SDS sample buffer either with or without reducing agent (mercaptoethanol), separated on a 10% SDS-PAGE and analyzed by Western blot using a polyclonal antibody to AP (GenHunter Corporation). Note the secreted 67 kDa AP alone does not form intermolecular disulfide bonds, whereas the secreted AP-T0 and AP-T2 fusions both are assembled efficiently into disulfide bond linked trimers.

B. Western blot analysis of the trimerization of soluble human TNF-RII when fused to the C-propeptides of α(I) collagen. The expression vectors encoding either the AP—C-propeptide fusion (T2) (as a negative control for antibody specificity), or human soluble TNF-RII-C-propeptide fusions as indicated in pTRIMER vectors were transiently

transfected into HEK293T cells. Forty-eight hours later, the conditioned media (20 μL) of each non-transfected and transfected cells as indicated were boiled for 5 minutes in equal volume of 2X SDS sample buffer either with or without reducing agent (mercaptoethanol), separated on a 10% SDS-PAGE and analyzed by Western blot using a monoclonal antibody to human TNF-RII (clone 226, R & D Systems, Inc.). Note the monoclonal antibody can only recognize the secreted TNF-RII with disulfide bonds. Both the soluble TNF-RII-T0 and TNF-RII-T2 fusions are assembled efficiently into disulfide bond linked trimers.

Fig.4. illustrates the bioassays showing the potent neutralizing activity of the trimeric soluble human TNF-RII-C-propeptide fusion protein against human TNF-α mediated apoptosis.

A. The TNF-α sensitive WEHI-13VAR cells (ATCC) were resuspended at 1 million cells/mL in RPMI medium containing 10% FBS. 100 μL of the cell suspension was plated into each well in a 96-well microtiter plate. Actinomycin D was added to each well at 500 ng/mL concentration followed by human TNF-α at 500 pg/ml (R & D Systems) in the presence or absence of trimeric soluble human TNF-RII-T2 as indicated. As a negative control, the trimeric AP-T2 was added in place of TNF-RII-T2. After 16 hours of incubation in a tissue culture incubator, the viability of cells was examined using either an inverted microscope at 20X magnification or cell viability indicator dye, Alamar Blue (BioSource, Inc.) added to 10% (v/v) to each well. The live cells are able to turn the dye color from blue to pink. Note that the trimeric soluble human TNF-RII-T2 exhibits a potent neutralizing activity against TNF-α □□□□□otects the cells from TNF-α mediated apoptosis

B. Quantitative analysis of the neutralizing activity of trimeric soluble human TNF-RII-T2 against human TNF-α. The experiment was carried out as Fig. 4A. Two hours after adding the Alamar Blue dye, the culture medium as indicted from each well was analyzed at OD575. The readings were normalized against wells with either no TNF-α (100% viability) added or with TNF-α without neutralizing agent (0% viability) added.

DESCRIPTION OF SEQUENCE LISTINGS

Sequence ID No. 1 (963 bases)

Nucleotide sequence encoding the C-propeptide human collagen α(I) T0 construct. The cDNA construct was cloned into the pAPtag2 vector, replacing the AP coding region. The underlined sequences denote restriction enzyme sites used in constructing the corresponding pTRIMER vector. The bolded codons denote the start and the stop of the T0 coding region.

Sequence ID No. 2 (311 aa)

The predicted C-propeptide T0 protein sequence of human Collagen α(I). The underlined sequence denotes the region of the "glycine repeats" upstream of the C-propeptide. The amino acid residues in red indicate the BMP-1 protease recognition site.

Sequence ID No. 3 (771 bases)

Nucleotide sequence encoding the C-propeptide of human collagen α(I) T2 construct. The cDNA construct was cloned into pAPtag2 vector, replacing the AP coding region. The underlined sequences denote restriction enzyme sites used in constructing the corresponding pTRIMER vector. The bolded codons denote the start and the stop of the T2 coding region.

Sequence ID No. 4 (247 aa)

The predicted C-propeptide T2 protein sequence of human Collagen α(I). The amino acid residue in red indicates the location of mutated BMP-1 protease recognition site.

Sequence ID No. 5 (2487 bases)

Nucleotide sequence encoding the human placental alkaline phosphatase (AP) fused to the T0 C-propeptide of human α(I) collagen (AP-T0). The underlined sequences indicate the restriction sites used for the fusion construct. The restriction site, which marks the fusion site shown in the middle of the sequence, is Bgl II.

Sequence ID No. 6 (819 aa)

The predicted protein sequence of the AP-T0 fusion protein. The amino acid residues in blue indicate fusion sites between human placental alkaline phosphates (AP) and the α(I) collagen T0 polypeptide. The bolded codons denote the start and the stop of the fusion protein. The underlined sequence denotes the region of the "glycine repeats" upstream of the C-propeptide of human α(I) collagen. The amino acid residues in red indicate the BMP-1 protease recognition sequence.

Sequence ID No. 7 (2294 bases)

Nucleotide sequence encoding the human placental alkaline phosphatass (AP) fused to the T2 C-propeptide human α(I) collagen (AP-T2). The bolded codons denote the start and the stop of the fusion protein. The underlined sequences indicate the restriction sites used for the fusion construct. The restriction site, which marks the fusion site shown in the middle of the sequence, is Bgl II.

Sequence ID No. 8 (755 aa)

The predicted protein sequence of the AP-T2 Fusion. The amino acid residues in blue indicate fusion sites between human placental alkaline phosphates (AP) and the α(I) collagen T2 polypeptide. The amino acid residue in red indicates the location of the mutated BMP-1 protease recognition site.

Sequence ID No. 9 (1734 bases)

Nucleotide sequence encoding the human soluble TNF-RII fused to the T0 C-propeptide of human α(I) collagen (sTNF-RII-T0). The bolded codons denote the start and the stop of the fusion protein. The underlined sequences indicate the restriction sites used for the fusion construct. The underlined sequence, which marks the fusion site shown in the middle of the sequence, is the BamH I/Bgl II ligated junction.

Sequence ID No. 10 (566 aa)

The predicted protein sequence of the human soluble TNF-RII-T0 Fusion. The amino acid residues in blue indicate fusion sites between human soluble TNF-RII and α(I) collagen T0 polypetide. The underlined sequence denotes region of the "glycine repeats" upstream of the C-propeptide of human α(I) collagen. The amino acid residues in red indicate the BMP-1 protease recognition site.

Sequence ID No. 11 (1542 bases)

Nucleotide sequence encoding the human soluble TNF-RII fused to the T2 C-propeptide of human α(I) collagen (sTNF-RII-T2). The bolded codons denote the start and the stop of the fusion protein. The underlined sequences indicate the restriction sites used for the fusion construct. The underlined sequence, which marks the fusion site shown in the middle of the sequence, is the BamH I/Bgl II ligated junction.

Sequence ID No. 12 (502 aa)

The predicted protein sequence of the human soluble TNF-RII-T2 fusion protein. The amino acid residues in blue indicate fusion sites between human soluble TNF-RII and the α(I) collagen T2 polypeptide. The amino acid residue in red indicates the location of mutated BMP-1 protease recognition site.

Sequence ID No. 13 (2139 bases)

Nucleotide sequence encoding the human soluble CD4 fused to the T0 C-propeptideof human α(I) collagen. The underlined sequences indicate the restriction sites used for the fusion construct. The underlined sequence, which marks the fusion site shown in the middle of the sequence, is the Bgl II site.

Sequence ID No. 14 (699 aa)

The predicted Protein Sequence of the human soluble CD4-T0 Fusion. The amino acid residues in blue indicate fusion sites between human soluble CD4 and α(I) collagen T0 polypeptide. The underlined sequence denotes the region of the "glycine repeats" upstream of the C-propeptide of human α(I) collagen. The amino acid residues in red indicate the BMP-1 protease recognition site.

Sequence ID No. 15 (1947 bases)

Nucleotide sequence encoding the human soluble CD4 fused to the T2 C-propeptide of human α(I) collagen. The underlined sequences indicate the restriction sites used for the fusion construct. The underlined sequence, which marks the fusion site shown in the middle of the sequence, is the Bgl II site.

Sequence ID No. 16 (635 aa)

The predicted Protein Sequence of the human soluble CD4-T2 Fusion. The amino acid residues in blue indicate fusion sites between human soluble CD4 and α(I) collagen T2

polypeptide. The amino acid residue in red indicates the location of mutated BMP-1 protease recognition site.

DESCRIPTION OF THE INVENTION

Prior to setting forth the invention, it may be helpful to an understanding thereof to set forth definitions of certain terms to be used hereinafter.

DNA Construct: A DNA molecule, generally in the form of a plasmid or viral vector, either single- or double-stranded that has been modified through recombinant DNA technology to contain segments of DNA joined in a manner that as a whole would not otherwise exist in nature. DNA constructs contain the information necessary to direct the expression and/or secretion of the encoding protein of interest.

Signal Peptide Sequence: A stretch of amino acid sequence that acts to direct the secretion of a mature polypeptide or protein from a cell. Signal peptides are characterized by a core of hydrophobic amino acids and are typically found at the amino termini of newly synthesized proteins to be secreted or anchored on the cell surface. The signal peptide is often cleaved from the mature protein during secretion. Such signal peptides contain processing sites that allow cleavage of the signal peptides from the mature proteins as it passes through the protein secretory pathway. A signal peptide sequence when linked to the amino terminus of another protein without a signal peptide can direct the secretion of the fused protein. Most of the secreted proteins, such as growth factors, peptide hormones, cytokines and membrane proteins, such as cell surface receptors, contain a signal peptide sequence when synthesized as a nascent protein.

Soluble receptor: The extracellular domain, in part or as a whole, of a cell surface receptor, which is capable of binding its ligand. Generally, it does not contain any

internal stretch of hydrophobic amino acid sequence responsible for membrane anchoring.

C-propeptide of collagens: The C-terminal globular, and non-triple-helical domain of collagens, which is capable of self-assembly into trimers. In contrast to the triple helical region of collagens, the C-propeptide does not contain any glycine repeat sequence and is normally proteolytically removed from procollagen precursor upon procollagen secretion before collagen fibril formation.

Glycine repeats: The central linear triple helix forming region of collagen which contains hundreds of (Gly-X-Y)n repeats in amino acid sequence. These repeats are also rich in proline at X or/and Y positions. Upon the removal of N-and C-propeptides, the glycine-repeats containing collagen triple helices can assemble into higher order of insoluble collagen fibrils, which makes up the main component of the cell matrix.

cDNA: Stands for complementary DNA or DNA sequence complementary to messenger RNA. In general cDNA sequences do not contain any intron (non-protein coding) sequences.

Prior to this invention, nearly all therapeutic antibodies and soluble receptor-Fc fusion proteins, such as Enbrel, are dimeric in structure (Fig. 1). Although these molecules, compared to their monomeric counterparts, have been shown to bind their target antigens or ligands with increased avidity, it is predicted that they are still imperfect, due to structural constrains, to bind their targets that have a homotrimeric structure. Examples of such therapeutically important trimeric ligands include TNF family of cytokines and HIV coat protein gp120. Therefore, from a structural point of view, it will be desirable to be also able to generate trimeric soluble receptors or

antibodies, which can perfectly dock to their target trimeric ligands or antigens (Fig. 1), and thereby completely block the ligand actions. Such trimeric soluble receptors or chimeric antibodies are expected to have the highest affinity to their targets and thus can be used more effectively and efficiently to treat diseases such as arthritis and AIDS.

This invention discloses ways for generating such secreted trimeric receptors and biological active proteins by fusing them to the C-propeptides of collagen, which are capable of self-assembly into trimers. The following are the advantages of this invention: (1) collagen is the most abundant protein secreted in the body of a mammal, constituting nearly 25% of the total protein in the body; (2) the major forms of collagen naturally occur as trimeric helixes, with their globular C-propetides responsible for the initiating of trimerization, which are subsequently proteolytically cleaved upon triple helix formation; (3) the cleaved soluble trimeric C-propeptide of collagen is found naturally at sub microgram/mL level in the blood of mammals; (4) the linear triple helical region of collagen can be included as a linker or excluded as part of the fusion protein so the distance between a protein to be trimerized and the C-propeptide of collagen can be precisely adjusted to achieve an optimal biological activity; (5) the recognition site of BMP1 which cleaves the C-propeptide off the pro-collagen can be mutated or deleted to prevent the disruption of a trimeric fusion protein; (6) the C-proptide domain provides a universal affinity tag, which can be used for purification of any secreted fusion proteins created by this invention; (7) unlike the IgG1 Fc tag which is known to be have other biological functions such as binding to its own cell surface receptors, the only known biological function of the C-propeptide of collagen is its ability to initiate trimerization of nascent pro-collagen chains and keep the newly made pro-collagen trimer soluble before

assembly into insoluble cell matrix. These unique properties of the C-propeptide of collagen would predict that this unique trimerization tag is unlikely going to be toxic, or immunogenic, making it an ideal candidate for therapeutic applications.

To demonstrate the feasibility for making secreted trimeric fusion proteins, cDNA sequences encoding the entire C-propeptides of human α1 (I) containing either some glycine-repeat triple helical region (T0 construct, sequence ID No. 1-2), or no glycine-repeat with a mutated BMP-1 recognition site (T2 construct, Sequence ID No. 3-4) were amplified by RT-PCR using EST clones purchased from the American Type Culture Collection (ATCC). The amplified cDNAs were each cloned as a Bgl II-XbaI fragment into the pAPtag2 mammalian expression vector (GenHunter Corporation; Leder et al., 1996 and 1998), replacing the AP coding region (Fig. 2). The resulting vectors are called pTRIMER, versions T2 and T0, respectively. The vectors allow convenient in-frame fusion of any cDNA template encoding a soluble receptor or biologically active protein at the unique Hind III and Bgl II sites. Such fusion proteins have the collagen trimerization tags located at the C termini, similar to native pro-collagens.

Example 1:

To demonstrate the feasibility of this invention, a cDNA encoding the human secreted placental alkaline phosphatase (AP), including its native signal peptide sequence, was cut out as a Hind III-Bgl II fragment from the pAPtag4 vector (GenHunter Corporation; Leder et al., 1996 and 1998) and cloned into the corresponding sites of the pTRIMER-T0 and pTRIMER-T2 vectors. The resulting AP-collagen fusion constructs (sequence ID No. 5-8) were expressed in HEK293T cells (GenHunter Corporation) after transfection. The successful secretion of the AP-collagen fusion proteins can be readily

determined by AP activity assay using the conditioned media of the transfected cells. The AP activity reached about 1 unit/mL (or equivalent to about 1 μg/mL of the fusion protein) 2 days following the transfection. To obtain HEK293T cells stably expressing the fusion proteins, stable clones were selected following co-transfection with a puromycine-resistant vector, pBabe-Puro (GenHunter Corporation). Clones expressing AP activity were expanded and saved for long-term production of the fusion proteins.

To determine if the AP-collagen fusion proteins are assembled into disulfide bond-linked trimers, conditioned media containing either AP alone or AP-T0 and AP-T2 fusions were boiled in SDS sample buffers containing either without (non-reducing) or with β-mercaptoethanol (reducing), separated by an SDS PAGE and analyzed by Western blot using an anti-AP polycloning antibody (GenHunter Corporation). AP alone without fusion exhibited as a 67 kDa band under both non-reducing and reducing conditions, consistent with the lack of any inter-molecular disulfide bonds as expected (Fig. 3A). In contrast, both AP-T0 and AP-T2 fusion proteins secreted were shown to be three times as big (about 300 kDa) under the non-reducing condition as those under the reducing condition (90-100 kDa), indicating that both fusion proteins were assembled completely into homotrimers (Fig. 3A). This result essentially reduces the concept of this invention to practice.

Example 2:

To provide a proof that new and therapeutically beneficial biological functions can be endowed to a trimeric fusion protein, next a trimeric human soluble TNF-RII (p75) receptor using a corresponding EST clone purchased from the ATCC was constructed. As described in Example 1, the N-terminal region of human TNF-RII,

including the entire ligand-binding region, but excluding the trans-membrane domain, was cloned in-frame, as a Bam H I fragment, into the Bgl II site of both pTRIMER-T0 and pTRIMER-T2 vectors (Sequence ID Nos. 9-12). The resulting fusion constructs were expressed in HEK293T cells following transfection. Stable clones were obtained by puromycine co-selection as described in Example 1. Western blot analysis under both non-reducing and reducing conditions was carried out to determine if the resulting soluble TNF-RII-collagen fusion proteins were indeed expressed, secreted and assembled into trimeric forms. As expected, the monoclonal antibody against human TNF-RII (clone 226 from R & D Systems, Inc.) clearly recognized the trimeric soluble TNF-fusion proteins expressed by both T0 and T2 fusion vectors as 220-240 kDa bands, which are about three times bigger than the corresponding monomeric fusion proteins (Fig. 3B). The TNF-RII antibody failed to detect monomeric fusion proteins under reducing conditions, consistent with the property specified by the antibody manufacturer. As a negative control for antibody specificity, neither the HEK293T cell alone, nor the cells expressing AP-T2 fusion protein expressed any TNF-RII (Fig. 3B).

To determine if the trimeric soluble TNF-RII receptors are potent inhibitors of its trimeric ligand TNF- α, TNF-α bioassay was carried out using a cytokine sensitive cell line WEHI-13VAR (ATCC) essentially as described previously (Mohler et al., 1993). The result shown in Fig. 4 clearly indicated that the trimeric soluble TNF-RII-C-propeptide fusion proteins are extremely potent in neutralizing the TNF-α mediated apoptosis of WEHI-13VAR cells in the presence of Actinomycin D (500 ng/mL) (Sigma). When human TNF-α (R & D Systems) was used at 0.5 ng/mL, the trimeric soluble TNF-RII-T2 (both from serum-free media or in purified form) had an apparent

Ki-50 (50% inhibition) of about 2 ng/mL or 8 X 10^{-12} M (assuming the MW of 240 kDa as homotrimer). This affinity to TNF-α is 4 orders of magnitude higher than that of the monomeric TNF-RII and at least 10-100 times higher than that of the dimeric soluble TNF-RII-Fc fusion, such as Enbrel (Mohler et al., 1993).

This crucial example proves that this invention can create trimeric fusion proteins with new biological properties that may have great therapeutic applications. Such soluble trimeric human TNF receptors may prove to be much more effective than the current dimeric soluble TNF receptor (e.g. Enbrel) on the market in treating autoimmune diseases such as RA. The dramatically increased potency of trimeric-TNF receptors could greatly reduce the amount of TNF blockers to be injected weekly for each patient, while improving the treatment and significantly lowering the cost for the patients. The improved potency of trimeric TNF receptors should also alleviate the current bottleneck in dimeric TNF receptor production, which currently can only meet the demands in treating about 100,000 patients in the United States.

Example 3.

The HIV virus, the cause of AIDS, infects and destructs primarily a special lineage of T lymphocytes in our body. These so called CD4+ T cells express a cell surface protein dubbed CD4, which is the receptor of HIV. HIV recognizes the CD4+ cells with its viral coat protein gp120 that binds to CD4. Notably, the gp120 exists as a giant homotrimeric complex on the viral surface, whereas the CD4 is monomeric on the cell surface. The current model for HIV infection is that of a complete docking of HIV to CD4+ T cells, when all three subunits of gp120 trimers are each bound to CD4 is required for viral RNA entry into the cells. Obviously, one of the straightforward strategies for stopping HIV

infection is to use soluble CD4 to blind the virus. Indeed, such approach using both monomeric soluble CD4 and CD4-Fc fusions has been shown quite effective in curbing HIV infections of laboratory isolates (Clapham et al., 1989; Daar et al., 1990). Unfortunately, these soluble CD4 were less effective in stopping the infection of HIV viral strains found in AIDS patients (Daar et al., 1990), possibly due to the amino acid sequence variations of the gp120, which lowers the affinity to monomeric and dimeric soluble CD4s.

To significantly increase the affinity of a soluble CD4 to any gp120 variants on HIV viruses, ideally a soluble CD4 should be in trimeric form so it can perfectly dock to its trimeric ligand, gp120 homotrimers. One of the major challenges for combating AIDS has been the high mutational rate of the viral genome, which leads to drug resistance. Therefore any drugs that directly target viral genes, such as HIV reverse transcriptase (e.g. AZT) and protease, are likely rendered ineffective as a result of viral mutations. In contrast, no matter how much it mutates, a HIV virus has to bind to a cellular CD4 receptor to initiate the infection. Thus, a high affinity soluble CD4 trimer should be immune to viral mutations because viral mutations in gp120 genes will render the virus unable to bind not only to a trimeric soluble CD4, but also CD4 on the cells.

To create such trimeric soluble CD4 HIV receptor analogs, a cDNA encoding the entire human soluble CD4, including its native signal peptide sequence, but excluding the transmembrane and the short cytoplasmic domains, was amplified using an EST clone purchased from the ATCC. The resulting cDNA was then cloned as a Hind III-Bgl II fragment into the corresponding sites of the pTRIMER-T0 and pTRIMER-T2 expression vectors. The resulting soluble CD4-collagen fusion constructs (sequence ID No. 13-16)

were expressed in HEK293T cells (GenHunter Corporation) after transfection. To obtain HEK293T cells stably expressing the fusion proteins, stable clones were selected following co-transfection with a puromycine-resistant vector, pBabe-Puro (GenHunter Corporation). Clones expressing the fusion proteins were expanded and saved for long-term production of the fusion proteins.

To determine if the souble human CD4-collagen fusion proteins are assembled into disulfide bond-linked trimers, conditioned media containing soluble CD4-T0 and CD4-T2 fusions were boiled in SDS sample buffers containing either without (non-reducing) or with β-mercaptoethanol (reducing), separated by a SDS PAGE and analyzed by Western blot using an monoclonal antibody to human CD4 (R & D Systems). Both soluble CD4-T0 and CD4-T2 fusion proteins secreted were shown to be three times as big (about 300 kDa) under the non-reducing condition as those under the reducing condition (90-100 kDa), indicating they were assembled essentially completely into homotrimers (data not shown). Now these trimeric soluble CD4 can be readily tested for gp120 binding and anti-HIV infection.